

Title: A method for determining a mimotope sequence

The invention relates to a method for determining a mimotope sequence.

Nowadays, random diversity libraries are widely used to identify lead molecules for diagnostics, pharmaca and 5 vaccines. When the lead molecule is a peptide, the methods used for identifying the lead molecules are often referred to as pepscan methods.

Pepscan methods have been known since the early eighties. The basic theory behind these methods is described 10 in EP-A-0 138 855 and EP-A-0 190 205.

Two pivotal aspects of screening of random diversity libraries are their size, i.e. how many different compounds are required to identify a lead molecule, and the method of optimization of the structure or sequence of the lead 15 molecule. Most methods described in the literature seem to be based on the idea that the larger the random diversity library, the higher the chance of finding a good lead molecule.

The aspect with respect to which the known methods 20 are most divergent, is the method of optimization of the lead molecule. Once the random diversity library has been synthesized, it can be tested for its desired activity. Of course it is possible to simply choose as a lead molecule the member of the random diversity library that score's the 25 highest value for said desired activity. However, most methods comprise steps for optimizing the structure of said member that shows the highest activity, in order to arrive at a lead molecule that shows an even higher activity.

Minipepscan libraries, composed of only a few 30 thousand peptides, have been used to identify lead molecules which were quickly optimized to molecules having an activity similar to that of molecules arrived at from starting from libraries composed of millions of compounds. This has been disclosed in Slootstra et al., Molecular Diversity, 1 (1995b) 87-96, and in Slootstra et al., J. Mol. Recogn., 10, 217-224 35

(1997). In these articles, it has been shown that small libraries in combination with pepscan-based optimization methods are a valuable tool in identifying and optimizing lead molecules.

5 The present invention aims to provide an improved method for identifying and optimizing a lead molecule.

It has been found that the objective improvement is achieved by starting from a library of known test molecules, evaluating the activity of said library, and by optimizing 10 the structure of a few of test molecules showing the highest activity, by methodically replacing each building block of the structure of the test molecules with all other possible building blocks.

Thus, the invention relates to a method for 15 determining a mimotope sequence for a receptor comprising the steps of:

- a) providing a library of test sequences;
- b) determining the activity of each test sequence of the library towards the receptor;
- 20 c) identifying a test sequence comprising at a certain position a building block which, according to the results of step b), is favored at said position;
- d) providing a next library of test sequences, based on said test sequence identified in step c), by replacing a 25 building block at selected positions of the identified test sequence with selected building blocks;
- e) determining the activity of each test sequence of the library provided in step d) towards the receptor;
- f) identifying a test sequence comprising at a certain 30 position a building block which, according to the results of step e), is favored at said position;
- g) repeating steps d) - f) for the library of test sequences provided in step d), for a number of cycles sufficient for finding in step f) a mimotope sequence that gives 35 sufficient activity towards the receptor.

It has been found that the present method leads, in a convenient manner, to a mimotope sequence having a very high binding strength to the desired receptor. In some cases, where the exact epitope (i.e. the partial sequence of an antigen, that provides binding to an antibody) is known, the present method has been found to lead to the exact structure of the epitope in only a few steps.

According to the invention, a mimotope sequence for a receptor is determined. In this context, a mimotope sequence is defined as a molecule that shows a certain, minimal, desired activity in the presence of a given receptor. An example hereof is the determination of the epitope sequence of an antigen for a certain antibody. However, in some cases it may be sufficient if the exact epitope is mimicked and that a slightly less binding strength of the found molecule to the receptor suffices.

The mimotope sequence will be a molecule that is composed of a number of building blocks, wherein the number and order of building blocks controls the properties of the molecule. Examples of types of mimotope sequences that may be determined according to the invention include peptides, possibly having steroid or saccharide-like structures connected thereto, saccharides, DNA (oligonucleotides), and PNA (peptide-like nucleic acids). Thus, the building blocks will be chosen from the groups of amino acids (both natural and non-natural amino acids), monosaccharides, and nucleotides.

The receptor for which a mimotope sequences is determined according to the invention may be any compound, composition, microorganism, or tissue sample towards which one of the types of mimotope sequences may show some activity that can be measured, such as a binding activity. Suitable examples of receptors include antibodies (both monoclonal and polyclonal), proteins, such as enzymes, cells, hormone receptors, and micro-organisms.

The first step in a method according to the invention is the provision of a library of test sequences. Of course, these test sequences are composed of the same type or class of building blocks as the objective mimotope sequence.

5 Preferably, the test sequences are known in that, although they may be randomly chosen, their composition and structure is known. This may be accomplished by generating, for instance by hand or by computer, a number of sequences and synthesizing the sequences thus obtained. It is also possible 10 to derive the library of test sequences from a compound which is known to have a favorable activity towards the receptor. In principle, any known manner of selecting test sequences to make up the library is suitable.

15 The length of the test sequences is dependent on the nature of the building blocks constituting the sequences, and on the nature of the receptor and the desired activity towards said receptor. For most purposes, a length of between 3 and 20 building blocks will be suitable.

20 The number of test sequences constituting the library will be large enough to provide sufficient data to come to a good mimotope sequence in an acceptable number of steps/cycles. On the other hand, said number will be small enough to assure that the data obtained can be handled quite conveniently. Usually, the number of test sequences in the 25 library will lie between 500 and 100,000, preferably between 1,000 and 10,000.

30 After the desired number of test sequences is generated, said test sequences may be synthesized. This may be done in any known manner, for instance as has been described by Slootstra et al. in Molecular Diversity, 1 (1995b), 87-96. Of course, it is also possible to make use of a library which is already available, for instance because it has been used in a previous run of the present method. In that case, the test sequences will not have to be 35 synthesized.

For reasons of convenience, the test sequences are preferably synthesized using a minicard or a flat support medium as has been described in Dutch patent application 10.019703.

5 The next step of a method according to the invention is the determination of the activity of each test sequence of the library towards the receptor. The manner wherein this determination is carried out will depend on the specific interaction between mimotope sequence and receptor that is 10 aimed at, and on the nature of the receptor and the building blocks of the mimotope sequence. For instance, when the desired activity is a binding of the mimotope sequence to the receptor, and the mimotope sequence is a peptide and the receptor is a monoclonal antibody, the determination may 15 suitable be performed in an ELISA test, either in solution or on a solid support. Other suitable methods of determining the activity include BIACORE and AFM (Atomic Force Microscope). The skilled person will be able to choose a suitable manner 20 of determination of the activity, given a certain receptor and nature of the mimotope sequence.

From the results of the determination of the activities of the test sequences of the library, at least one test sequence will be chosen to form a basis for the remaining steps of the present method. Said at least one test 25 sequence is, in accordance with the invention, chosen by identifying at a certain position a building block which, as appears from the results of the determination of the activity, is favored at said position. In this regard, by the phrase "at a certain position a building block which is 30 favored at said position" is meant that test sequences having at said position said building block show a high activity towards the receptor, relative to other activities found.

It is possible that the at least one test sequence which is identified was tested for its activity towards the 35 receptor and showed itself a high activity, relative to other activities found.

It is also possible that the at least one test sequence which is identified, was in itself not tested. This can be explained as follows. For instance, it may have been found that a test sequence having at position 2 building block A shows a very high activity towards the receptor. It may further have been found that a test sequence having at position 4 building block B also shows a very high activity towards the receptor. The at least one test sequence to be identified may then be a sequence comprising building block A at position 2 and building block B at position 4. However, it may very well be that this sequence itself had not yet been evaluated for its activity towards the receptor.

Generally, the number of test sequences identified for basing the remaining steps of a method according to the invention on will be between 1 and 150. A higher number of test sequences chosen will lead to better results, but will be relatively more cumbersome to handle. Preferably, said number will be between 1 and 25. The skilled person will be able to decide on the number of test sequences chosen, dependent on the desired quality of the result of the method and the facilities available for carrying out the method.

On the basis of the identified test sequences, a next library of test sequences is provided in a so-called replacement analysis. This is done by varying the building blocks at positions of the identified test sequences, i.e. by replacing a building block at selected positions of the identified test sequence with selected building blocks.

For instance, in its simplest form the replacement analysis is carried out as follows. In case the test sequences are dodecapeptides of natural amino acids, first, the building block at position one of one of the chosen dodecapeptides may changed into each of the twenty naturally occurring amino acids, leading to 20 new test sequences. This is repeated for each position of the dodecapeptide, leading to $20 * 12 = 240$ test sequences. Among these 240 sequences, the original dodecapeptide will be present 12 times. This

procedure is repeated for each chosen dodecapeptide. Thus, a next library is obtained which comprises from 240 to 3600 new test sequences.

More complex forms of this replacement analysis may 5 be performed, for instance by allowing certain building blocks to be replaced by groups of building blocks (e.g. the replacement of one amino acid by two or more amino acids). Usually, such a replacement by multiple building blocks will 10 be used sparingly. Preferably, in at least one replacement analysis of the different cycles of the present method, at least one building block is replaced by a group of building blocks. Care should be taken that in each replacement analysis not too many multiple building blocks are introduced, for then problems concerning the length of the 15 sequence may arise. For the same reason, the multiple group of building blocks should preferably not be too long. The skilled person will, based on his experience in the field, be able to judge which size and number of multiple building blocks may be introduced.

It is also possible to replace a building block by a void in the replacement analysis. That way, the sequence will 20 become one building block shorter. It will be clear that not too many building blocks should be replaced by a void at the same time, as this could lead to an undesirably short (or no) 25 test sequence. The skilled person will be able to judge when and at which position it is helpful to introduce a void into a test sequence.

In addition, it is possible to allow the introduction 30 of selected building blocks in the replacement analysis only, or to only replace building blocks at selected positions in the test sequences. Sometimes it is preferred to maintain certain building blocks at certain positions, e.g. in order to maintain a desired three dimensional structure of the test sequence. An advantage of this manner of performing the 35 replacement analysis is that the total number of test sequences to be provided is limited. Generally, however, if

the building blocks at all positions of the test sequence are replaced by all possible building blocks, the test sequences having the desired three dimensional structure will also be provided. Again, it is within the skill of the artisan to 5 judge in which cases it may be helpful to carry out a replacement at selected positions only. He will also be able to judge to which positions this may apply, and which building blocks may be used for the replacement and which building blocks may best not be used in the replacement (at 10 certain positions).

The test sequences of the next library, which provided on the basis of the replacement analysis of the at least one identified test sequence, are tested for their activity towards the receptor. In case a test sequence is 15 present among the test sequences of this library that shows sufficient activity towards the receptor, the method is completed and the desired mimotope sequence is obtained. It will be clear that it will depend on the type of activity, the receptor, the mimotope sequence, and the objective 20 application of the mimotope sequence, whether a certain activity is regarded sufficient. Given the circumstances, the skilled person will be able to choose a threshold for a desired activity.

In case none of the test sequences shows sufficient 25 activity towards the receptor, another cycle is carried out. In accordance with the invention, preferably at least two cycles, and thus two replacement analyses, are carried out. Thus, a next library of test sequences is provided, based on the 10 to 15 test sequences of the previous library that 30 showed the highest activity, as set forth above. The test sequences in this next library will be evaluated for their activity, and so on.

The number of cycles that has to be carried out will depend on the desired degree of activity of the mimotope 35 sequence to be found for the receptor. Starting from about 4500 randomly chosen dodecapeptides, it has been found

possible to arrive at the exact epitope sequence for a monoclonal antibody in less than three cycles.

In a preferred embodiment of the invention, the amount of receptor used for determining the activity of the test sequences towards said receptor, is lowered in each cycle. It has been found that building blocks that seem essential in the first cycle may become non-essential in further cycles. In other words, in said first cycle local optima may be found, which may be overcome in a subsequent cycle or in subsequent cycles. The lowering of the amount of the receptor in the determining of the activity of the test sequences has been found to assist in overcoming such local optima. Preferably, the amount of receptor used in the determination of the activities of test sequences in a library is reduced by a factor of from 50 to 1000, more preferably from 10 to 100.

The invention will now be further elucidated by the following, non-restrictive examples.

20

EXAMPLES

The described examples illustrate six variations of the present optimization method that has been successfully applied to lead peptides, identified with antibodies (monoclonal and polyclonal) raised against peptides, proteins, viruses, bacteria, sugars and steroids, from various types of peptide libraries. It can be envisaged that additional variations are possible. The concept that rules these variations is the repetition of replacement analyses on 1 or more lead peptides. Local minima are overcome and epitopes and mimotopes are identified.

Materials and Methods**Synthesis and screening minicards libraries (used in examples 1, 2, 5, 6)**

5 Using the 20 natural L-amino acids 4550 random dodecapeptides (12-mers) were generated with a random generator programmed in Quick basic which runs on a 486 DX2 (66 MHz) computer system. In this library the frequency of each residue is approximately 5%. This set of sequences was
10 used to design additional libraries. One in which all sequences are in the D-amino acid form, one in which the 2nd and 11th position are held by a cysteine and one in which the 3rd and 10th position are held by a cysteine. The aim of the latter two is to introduce a disulphide bridge into the
15 peptide which should present the peptides as loops. Very important here is that the sequence of the amino acids within and outside the cysteines is identical to that of the first set of 4550 dodecapeptides. In this way the activity of these sets of dodecapeptides can be compared. **Example-1** describes
20 results obtained with the synthetic peptide library one in which the 3rd and 10th position are held by a cysteine. **Example-2** describes results obtained with the synthetic peptide library composed of 4550 random dodecapeptides (12-mers), i.e. no cysteine or any other motif was used.

25 The libraries were synthesized and screened using credit-card format mini-PEPSCAN cards (455 peptides/card) as described previously (Slootstra et al., 1995b). In **example-1** the binding of monoclonal antibody 26/9 to each peptide was tested in a PEPSCAN-based enzyme-linked immuno assay (ELISA).
30 Monoclonal antibody 26/9 has been described previously (Wilson et al., 1984; Rini et al., 1992; Churchill et al., 1994). Monoclonal antibody 26/9 was raised against the peptide HA175-110 of the hemagglutinin protein of influenza virus (X47:HA1) (Wilson et al., 1984).

35 In **example-1** the 455-well creditcard-format polyethylene cards, containing the covalently linked

peptides, were incubated with antibody 26/9 (100 μ g/ml). After washing the peptides were incubated with rabbit-anti-mouse peroxidase (rampo, dilution 1/1000) (Dakopatts) (1 hr, 25°C), and subsequently, after washing the peroxidase substrate 2,2'-azino-di-3-ethylbenzthiazoline sulfonate (ABTS) and 2 μ l/ml 3% H₂O₂ were added. After 1 hr the color development was measured. The color development of the ELISA was quantified with a CCD-camera and an image processing system. The setup consists of a CCD-camera and a 55 mm lens (Sony CCD Video Camara XC-77RR, Nikon micro-nikkor 55 mm f/2.8 lens), a camara adaptor (Sony Camara adaptor DC-77RR) and the Image Processing Software package TIM, version 3.36 (Difa Measuring Systems, The Netherlands). TIM runs on a 486 DX2 (50 MHz) computer system.

The dodecapeptide library, composed of 4550 random dodecapeptides, is screened with relatively high concentration of antibody (100 μ g/ml, more usually 10 μ g/ml, see examples 2-6). This concentration is approximately 2 orders of magnitude above the concentration required to obtain maximal binding activity in ELISA with the native epitope peptide (not shown). The concentration of 100 μ g/ml was used to obtain many binding peptides from the small set of 4550 random dodecapeptides. In previous studies it was shown that such a strategy, without obtaining an unfavorable signal to noise ratio, can result in hundreds of binding peptides that resemble small linear or non-linear parts of the native epitope (Slootstra et al., 1995b; Slootstra et al., 1997).

Synthesis and screening of OTHER pepscan libraries (used in example 4)

In addition to random minipepscan libraries mimotopes can also be identified from standard pepscan libraries. These libraries contain all overlapping 12-mers (or shorter/longer) covering the linear sequence of a known protein (Slootstra et al., 1995a).

Synthesis and screening of non-pepscan libraries (used in examples 3, 6)

Lead peptides can also be derived from other type of libraries such as for example phage-display libraries.

5

Sequence analysis of lead molecules (used in example-5)

All 4550 dodecapeptides were ranked according to their binding activity. According to their ranking consensus sequences and motifs are identified. Initial methods of lead 10 optimization that only use the sequence of the top 50 molecules have been described in Slootstra et al. (1995b; 1997). The present method uses all 4550 sequences. Firstly,

the frequency and/or distribution of single amino acids and the dipeptide motifs O0, dipeptide motifs OXO and dipeptide 15 motifs OXXO is determined using Microsoft Excel 4.0 (O, one of the 20 natural L-amino acids; X, any residue). Secondly, properties of the amino acids (hydrophobicity, charge etc.) are included in the analysis. Thirdly, all this data is used to optimize the activity of the top 10-50 lead molecules.

20 This is done by substitution of building blocks that inhibit activity and include building blocks that improve binding activity.

25 The present sequence analysis method improves the activity of lead molecules through motif analysis of all sequences part of the library.

Replacement analyses (used in examples 1, 2, 3, 4, 5, 6)

In a replacement analysis the binding activity of complete series of substitutions analogs of lead peptides, in 30 which each position is replaced by each of the other 19 natural L-amino acids, is investigated in detail. In addition, it is possible to use the 20 D-amino acids as well as other available non-natural amino acids. Substitutions that result in improved binding are combined into new 35 sequences of building blocks. These new sequences are again tested in a 2nd replacement analysis. If necessary more

rounds of replacement analyses are performed. After each round of a replacement analyses the antibody (or other soluble receptor) concentration can be lowered to obtain maximal binding activity (e.g. from 100 µg/ml to 0.01 µg/ml after 2 rounds of replacement-analyses, as shown in example-5 1).

References

- Churchill M.E.A., Stura E.A., Pinilla C., Appel J.R., Houghten R.A., Kono D.H., Balderas R.S., Fieser G.G., Schulze-Gahmen U. and Wilson I.A. (1994). Crystal structure of a peptide complex of anti-influenza peptide antibody Fab 26/9. *J. Mol. Biol.* 241, 534-556.
- Rini J.M., Schulze-Gahmen U. and Wilson I.A. (1992). Structural evidence for induced fit as a mechanism for antibody-antigen interaction. *Science*, 255, 959-965.
- Slootstra J.W., De Geus P., Haas H. Verrrips, C.T. and Meloen R.H. (1995a). Possible active site of the sweet-tasting protein thaumatin. *Chemical Senses*, 20, 535-543.
- Slootstra J.W., Puijk W.C. Ligtvoet G.J., Langeveld J.P.M. and Meloen R.H. (1995b). Structural aspects of antibody-antigen interaction revealed through small random peptide libraries. *Molecular Diversity* 1, 87-96.
- Slootstra J.W., Kuperus D., Plückthun, A. and Meloen R.H. (1996). Identification of new tag sequences with differential and selective recognition properties for the anti-FLAG monoclonal antibodies M1, M2 and M5. *Molecular Diversity* 2, 156-164.
- Slootstra J.W., Puijk W.C. Ligtvoet G.J., Kuperus D., W.M.M. Schaaper and Meloen R.H. (1997). Screening of a small set of random peptides: A new strategy to identify peptides that mimic epitopes. *J. Mol. Recog.* 10, 219-224.
- Wilson I.A., Niman H.L., Houghten R.A., Cherenson A.R., Connolly M.L. and Lerner R.A. (1984). The structure of 35 an antigenic determinant in a protein. *Cell*, 37, 767-778.

Notes for the examples:

- the 10 μ g/ml, 1 μ g/ml etc. denote the antibody concentration.

5 - The line "improved position-01" etc. means that this change in amino acid (is underlined) improves binding activity in pepscan elisa at given antibody concentration. Positions that could not be improved are not mentioned (e.g. position-08 in first replacement-net of example-1, i.e. the original I at position 8 gives the highest activity).

10

Example-1:

15 **title:** Identification of an epitope through a lead peptide selected from 4550 random dodecapeptides (minipepscan library with motif XXCXXXXXXCXX, X , randomly selected amino acid, cf. Materials and Method))

20

methodology: two rounds of replacement analyses

Results:

25 In example-1, the lead peptide CGCAAMNIRCYA was derived (with antibody 26/9) from the random library in which the 3rd and 10th position are held by a cysteine (see above in Materials and methods). All possible single substitution analogs of the random peptide CGCAAMNIRCYA tested in pepscan for binding antibody 26/9 at 100 μ g/ml. Below these results 30 are described in detail.

35 The replacement analysis of the lead peptide CGCAAMNIRCYA resulted in the identification of building blocks that cannot be replaced by any other building blocks (e.g. Y and A), and to the identification of residues that can be replaced by one or two other building blocks (e.g. N), and to residues that can be replaced by many other building

blocks (e.g. M). Some replacements improve binding activity considerably (e.g. N into D).

5 All these replacements were used to design the improved peptide EMDEEEDIMNYA. Note that this peptide binds antibody 26/9 at much lower concentration, i.e. has improved binding activity (100 μ g/ml for CGCAAMNIRCYA and 1.0 μ g/ml for EMDEEEDIMNYA).

10 The peptide EMDEEEDIMNYA was run through a second replacement analysis. Again some replacements improve binding activity considerably. These replacements were used to design the improved peptide EMDEEEDVPDYA. Essential is that the first part of EMDEEEDIMNYA, EMDEEE, does not contain critical residues whereas the latter part, DIMNYA, does. Combination of the improved residues in this latter part results in the 15 sequence DVPDYA. The sequence DVPDYA is identical to the linear epitope of antibody 26/9. Thus, the lead peptide CGCAAMNIRCYA derived from a few thousand random dodecapeptides was turned into native epitope sequence through two replacement analyses.

20

REPLACEMENT ANALYSIS-I

original lead: CGCAAMNIRCYA

activity at \geq 100.0 μ g/ml

improved position-01: EGCAAMNIRCYA

activity at \geq 10.0 μ g/ml

improved position-02: CMCAAMNIRCYA

activity at \geq 10.0 μ g/ml

improved position-03: CGDAAMNIRCYA

activity at \geq 10.0 μ g/ml

30 improved position-04: CGCEAMNIRCYA

activity at \geq 10.0 μ g/ml

improved position-05: CGCAEMNIRCYA

activity at \geq 10.0 μ g/ml

improved position-06: CGCAAENIRCYA

activity at \geq 10.0 μ g/ml

35 improved position-07: CGCAAAMDIRCYA

activity at $\geq 10.0 \mu\text{g/ml}$
improved position-09: CGCAAMNIMCYA
activity at $\geq 10.0 \mu\text{g/ml}$
improved position-10: CGCAAMNIRNYA
activity at $\geq 10.0 \mu\text{g/ml}$
5 combination improvements: EMDEEEDIMNYA
activity at $\geq 1.0 \mu\text{g/ml}$

REPLACEMENT ANALYSIS-II

10 combinations rep-an. I: EMDEEEDIMNYA
activity at $\geq 1.0 \mu\text{g/ml}$
improved position-08: EMDEEEDVMNYA
activity at $\geq 0.1 \mu\text{g/ml}$
improved position-09: EMDEEEDIPNYA
activity at $\geq 0.1 \mu\text{g/ml}$
15 improved position-10: EMDEEEDIMDYA
activity at $\geq 0.1 \mu\text{g/ml}$
combination improvements: EMDEEEDVPDYA
activity at $\geq 0.01 \mu\text{g/ml}$

20 Notes:

The sequence DVPDYA is the original epitope. The peptide EMDEEEDVVPDYA has a 10-fold improved binding affinity (in solution) over the native epitope peptide YPYDVPDYASLRS.

25

Example-2:

title: Identification of a mimotope through a lead peptide
30 selected from 4550 random dodecapeptides (random minipepscan library)

tools: lead peptide ANWPSAIGAFGL

35 methodology: three rounds of replacement analyses

Results:

In example-2, lead peptides were identified through a random minipepscan library. The difference of example-2 with example-1 is that the antibody used in example-1 binds a linear epitope whereas the antibody used in example-2 binds a non-linear epitope. The following multiple replacement analyses were done as discussed in example-1. The identified mimotope does not resemble any region of the linear sequence of the native protein.

10

REPLACEMENT ANALYSIS-I

original lead:	ANWPSAIGAFGL
	activity at \geq 10.0 μ g/ml
improved position-01:	<u>H</u>NWPSAIGAFGL
	activity at \geq 5.0 μ g/ml
improved position-02:	<u>A</u>WWPSAIGAFGL
	activity at \geq 5.0 μ g/ml
improved position-03:	<u>A</u>NAPS AIGAFGL
	activity at \geq 5.0 μ g/ml
improved position-04:	ANW<u>S</u>SAIGAFGL
	activity at \geq 5.0 μ g/ml
improved position-11:	ANWSSAIGAF<u>KL</u>
	activity at \geq 5.0 μ g/ml
combination improvements:	<u>H</u>WASSAIGAF<u>KL</u>
	activity at \geq 1.0 μ g/ml

15

20

25

REPLACEMENT ANALYSIS-II

combinations rep-an. I:	HWASSAIGAFKL
	activity at \geq 1.0 μ g/ml
improved position-01:	<u>K</u>WASSAIGAFKL
	activity at \geq 0.5 μ g/ml
improved position-02:	<u>H</u>YASSAIGAFKL
	activity at \geq 0.5 μ g/ml
improved position-03:	<u>H</u>WGSSAIGAFKL
	activity at \geq 0.5 μ g/ml
improved position-07:	HWASSA<u>M</u>GAFKL

30

35

activity at $\geq 0.5 \mu\text{g/ml}$

combination improvements: KYGSSAMGAFKL

activity at $\geq 0.1 \mu\text{g/ml}$

5

REPLACEMENT ANALYSIS-III

combinations rep-an. II: KYGSSAMGAFKL

activity at $\geq 0.1 \mu\text{g/ml}$

improved position-03: KYFSSAMGAFKL

10

activity at $\geq 0.05 \mu\text{g/ml}$

improved position-06: KYGSSGMGAFKL

activity at $\geq 0.05 \mu\text{g/ml}$

combination improvements: KYFSSGMGAFKL

activity at $\geq 0.01 \mu\text{g/ml}$

15

Example-3:

title: Identification of a mimotope through a lead derived from phage-display library composed of >1000.000 random hexapeptides.

tools: lead peptide SDTRKG*

25 **methodology:** lengthening to the left and to the right of SDTRKG with cysteines/glycines, followed by three rounds of replacement-analyses

Results:

30 In example-3, lead peptides were identified through phage-display. Adjacent cysteines/glycines were added to improve binding activity. After obtained activity multiple replacement analyses were done as discussed in example-1.

LENGTHENING BY CYSTEINES/GLYCINES:

phage-display lead: SDTRKG
NO activity at 10.0 μ g/ml

lengthened lead: CSDTRKGCG
activity at \geq 10.0 μ g/ml

5 lengthened lead: CSDTRKGCG
activity at \geq 10.0 μ g/ml

REPLACEMENT ANALYSIS-I

10 lengthened lead: CSDTRKGCG
activity at \geq 10.0 μ g/ml

improved position-02: CTDTRKGCG
activity at \geq 5.0 μ g/ml

improved position-03: CSETRKGCG
activity at \geq 5.0 μ g/ml

15 improved position-05: CSDTHKGCG
activity at \geq 5.0 μ g/ml

improved position-06: CSDTRYGCG
activity at \geq 5.0 μ g/ml

20 combination improv.: CTETHYGCG
activity at \geq 1.0 μ g/ml

REPLACEMENT ANALYSIS-II

25 comb. rep-an. I: CTETHYGCG
activity at \geq 1.0 μ g/ml

improved position-02: CYETHYGCG
activity at \geq 0.5 μ g/ml

improved position-05: CTETKYGCG
activity at \geq 0.5 μ g/ml

30 improved position-06: CTETHFGCG
activity at \geq 0.5 μ g/ml

combination improv.: CYETKFGCG
activity at \geq 0.1 μ g/ml

REPLACEMENT ANALYSIS-III

comb. rep-an. II: CYETKFGCG
activity at $\geq 0.1 \mu\text{g/ml}$
improved position-01: DYETKFGCG
activity at $\geq 0.05 \mu\text{g/ml}$
5 improved position-08: CYETKFGNG
activity at $\geq 0.05 \mu\text{g/ml}$
combination improv.: DYETKFGNG
activity at $\geq 0.01 \mu\text{g/ml}$

10 *, this lead peptide is, as a synthetic peptide, not active (not in elisa (pepscan or standard nor in solution). This is not unique. Often phage-peptides are only active as part of the phage-coat protein. In other formats they lose their
15 activity.

Example-4:

20 **title:** Identification of a mimotope through a lead derived from a standard pepscan analysis.

tools: lead peptide RVMIKLILVNFR* and complete sequence of native protein (part which was used is KIYRVMIKLILVNFRMQP).

25 **methodology:** Two rounds of replacement-analyses, followed by lengthening to the left and right, again followed by two more rounds of replacement-analyses, finally followed by lengthening to an 18-mer mimotope.

30 **Results:**

In example-4, a lead peptide was identified through standard pepscan analysis, i.e. the antibody was tested on all overlapping 12-mers covering the linear sequence of the
35 protein. The following two rounds of replacement analyses were done as discussed in example-1. In addition, the

mimotope was lengthened to the left and right (3 amino acids, this can be further lengthened) through additional replacement analyses.

5 REPLACEMENT ANALYSIS-I

original lead: RVMIKLILVNFR
activity at \geq 10.0 μ g/ml

improved position-01: AVMIKLILVNFR
activity at \geq 5.0 μ g/ml

10 improved position-02: AIMIKLILVNFR
activity at \geq 5.0 μ g/ml

improved position-03: AVPIKLILVNFR
activity at \geq 5.0 μ g/ml

improved position-08: AVMIKLIRVNFR
activity at \geq 5.0 μ g/ml

15 improved position-11: AVMIKLILVNYR
activity at \geq 5.0 μ g/ml

combination improv.: AIPIKLIRVNYR
activity at \geq 1.0 μ g/ml

20

REPLACEMENT ANALYSIS-II

comb. rep-an. I: AIPPIKLIRVNYR
activity at \geq 1.0 μ g/ml

improved position-01: YIPIKLIRVNYR
activity at \geq 0.5 μ g/ml

25 improved position-02: APPIKLIRVNYR
activity at \geq 0.5 μ g/ml

combination improv.: YPPIKLIRVNYR
activity at \geq 0.1 μ g/ml

30

LENTHENING WITH NATIVE SEQUENCE:

comb. rep-an. II: YPPIKLIRVNYR
activity at \geq 0.1 μ g/ml

lengthened left: KIYYPPPIKLIRV
activity at \geq 1.0 μ g/ml

35 lengthened right: IKLIRVNYRMQP

activity at $\geq 1.0 \mu\text{g/ml}$

REPLACEMENT ANALYSIS-III of left lengthened peptide:

	lengthened left:	KIYYPPIKLIRV
5		activity at $\geq 1.0 \mu\text{g/ml}$
	improved posit-01:	<u>RIYYPPIKLIRV</u>
		activity at $\geq 0.5 \mu\text{g/ml}$
	improved posit-02:	K <u>P</u> YYPPIKLIRV
		activity at $\geq 0.5 \mu\text{g/ml}$
1	improved posit-03:	KI <u>W</u> YPPPIKLIRV
		activity at $\geq 0.5 \mu\text{g/ml}$
	improved posit-08:	KIYYPP <u>I</u> SLIRV
		activity at $\geq 0.5 \mu\text{g/ml}$
15	combination impr.:	<u>RPWYPPISLIRV</u>
		activity at $\geq 0.1 \mu\text{g/ml}$

REPLACEMENT ANALYSIS-IV of right lengthened peptide:

	lengthened right:	I <u>K</u> LIRVNYRMQP
		activity at $\geq 1.0 \mu\text{g/ml}$
20	improved posit-10:	I <u>K</u> LIRVNYRCQP
		activity at $\geq 0.5 \mu\text{g/ml}$
	improved posit-11:	I <u>K</u> LIRVNYRM <u>E</u> P
		activity at $\geq 0.5 \mu\text{g/ml}$
	improved posit-12:	I <u>K</u> LIRVNYRM <u>Q</u> N
25		activity at $\geq 0.5 \mu\text{g/ml}$
	combination impr.:	<u>I</u> KLIRVNYRC <u>E</u> N
		activity at $\geq 0.1 \mu\text{g/ml}$
	combination total:	RPWYPPISLIRVNYRCEN
30		activity at $\geq 0.01 \mu\text{g/ml}$

*, lead peptide was identified from a library composed of all overlapping 12-mers covering the linear sequence of a protein. This makes it possible to lengthen the peptide to the left and right with adjacent amino acids, in this case KIY on the left and MQP on the right. Additional replacement-

analyses of 12-mers, shifted three to the left or three to the right of the sequence finally results in an 18-mer mimotope.

5

Example-5:

10 **title:** Identification of a mimotope through a set of similar leads selected from 4550 random dodecapeptides (random minipepscan library)

15 **tools:** set of 6 lead peptides QNNMKLFRGCVP, RGIKWNEMTDQW, KLQQNPTFYPPV, TNNCKEFAGIVP, RGILTNIMKDQW, IVQNNPKFFRGA (potentially up to all 4550 peptides, see material and method, in this case also 'negative' amino acids are removed from the leads)

20 **methodology:** : determination of consensus sequence, followed by two rounds of replacement analyses of the consensus sequence

Results:

25 In example-5, a set of similar lead peptides were used to identify a consensus sequence. The consensus sequences were identified as discussed in materials and method. This sequence was used in the replacement analyses.

ALIGNMENT OF LEAD PEPTIDES:

30 QNNMKLFRGCVP
 RGIKWNEMTDQW
 KLQQNPTFYPPV
 TNNCKEFAGIVP
 RGILTNIMKDQW
35 IVQNNPKFFRGA

consensus: ILQNNMKDFRG

REPLACEMENT ANALYSIS-I

consensus lead:

ILQNNMKDFRG

activity at $\geq 1.0 \mu\text{g/ml}$

5 improved position-03:

ILTNNMKDFRG

activity at $\geq 0.5 \mu\text{g/ml}$

improved position-07:

ILQNNMPDFRG

activity at $\geq 0.5 \mu\text{g/ml}$

10 improved position-10:

ILQNNMKDWRG

activity at $\geq 0.5 \mu\text{g/ml}$

combination improv.:

ILTNNMPDWRG

activity at $\geq 0.1 \mu\text{g/ml}$

15 REPLACEMENT ANALYSIS-II

comb. rep-an. I:

ILTNNMPDWRG

activity at $\geq 0.1 \mu\text{g/ml}$

improved position-07:

ILTNNMGDWRRG

activity at $\geq 0.05 \mu\text{g/ml}$

20 improved position-11:

ILTNNMPDWYRG

activity at $\geq 0.05 \mu\text{g/ml}$

combination improv.:

ILTNNMGDWYRG

activity at $\geq 0.01 \mu\text{g/ml}$

25

Example-6:

title: Identification of a mimotope through a set of
different leads selected from phage-display library composed
30 of >1,000,000 random hexapeptides.

tools: set of 3 lead peptides (ANWPSA, KLITRW, NVCSWS)

methodology: Two rounds of replacement analyses (for each
35 lead peptide), followed by determination of overall consensus
sequence.

Results:

5 In example-6, different lead peptides were used to identify a consensus sequence. Each lead peptide was used in multiple replacement analyses (two rounds). The resulting three mimotopes were aligned which resulted in a lengthened mimotope with improved activity.

REPLACEMENT ANALYSIS-IA

original lead	:	ANWPSA	:	activity at \geq 10.0	$\mu\text{g}/\text{ml}$
10 improved position-01:	<u>H</u> NWPSA	:	activity at \geq 5.0	$\mu\text{g}/\text{ml}$	
improved position-02:	<u>A</u> WWPSA	:	activity at \geq 5.0	$\mu\text{g}/\text{ml}$	
improved position-03:	<u>A</u> NAPSA	:	activity at \geq 5.0	$\mu\text{g}/\text{ml}$	
improved position-04:	ANW <u>S</u> SA	:	activity at \geq 5.0	$\mu\text{g}/\text{ml}$	
combination improv.	<u>H</u> WA <u>S</u> SA	:	activity at \geq 1.0	$\mu\text{g}/\text{ml}$	

15

REPLACEMENT ANALYSIS-IIA

comb. rep-an. IA	:	HWASSA	:	activity at \geq 1.0	$\mu\text{g}/\text{ml}$
improved position-05:	HWASPA	:	activity at \geq 0.5	$\mu\text{g}/\text{ml}$	
combination improv.	<u>H</u> WASPA	:	activity at \geq 0.5	$\mu\text{g}/\text{ml}$	

20

REPLACEMENT ANALYSIS-IB

original lead	:	KLITRW	:	activity at \geq 10.0	$\mu\text{g}/\text{ml}$
improved position-01:	<u>S</u> LITRW	:	activity at \geq 5.0	$\mu\text{g}/\text{ml}$	
improved position-02:	<u>K</u> SITRW	:	activity at \geq 5.0	$\mu\text{g}/\text{ml}$	
25 improved position-03:	<u>K</u> LATRW	:	activity at \geq 5.0	$\mu\text{g}/\text{ml}$	
improved position-06:	KLIT <u>R</u>	:	activity at \geq 5.0	$\mu\text{g}/\text{ml}$	
combination improv.	<u>S</u> SATRY	:	activity at \geq 1.0	$\mu\text{g}/\text{ml}$	

25

REPLACEMENT ANALYSIS-IIB

30 comb. rep-an. IB	:	SSATRY	:	activity at \geq 1.0	$\mu\text{g}/\text{ml}$
improved position-02:	<u>S</u> PATRY	:	activity at \geq 0.5	$\mu\text{g}/\text{ml}$	
combination improv.	<u>S</u> PATRY	:	activity at \geq 0.5	$\mu\text{g}/\text{ml}$	

30

REPLACEMENT ANALYSIS-IC

35 original lead	:	NVCSWS	:	activity at \geq 10.0	$\mu\text{g}/\text{ml}$
improved position-02:	<u>N</u> ICWS	:	activity at \geq 5.0	$\mu\text{g}/\text{ml}$	

improved position-04: NVCHWS : activity at \geq 5.0 $\mu\text{g}/\text{ml}$
improved position-06: NVCSWA : activity at \geq 5.0 $\mu\text{g}/\text{ml}$
combination improv. : NICHWA : activity at \geq 1.0 $\mu\text{g}/\text{ml}$

5 REPLACEMENT ANALYSIS-IIC

comb. rep-an. IC : NICHWA : activity at \geq 1.0 $\mu\text{g}/\text{ml}$
 improved position-01: YICHWA : activity at \geq 0.5 $\mu\text{g}/\text{ml}$
 improved position-02: NVCHWA : activity at \geq 0.5 $\mu\text{g}/\text{ml}$
 combination improv. : YVCHWA : activity at \geq 0.1 $\mu\text{g}/\text{ml}$

10

ALIGNMENT combination improv. IIA, IIB and IIC:

combination improvements 1: HWASPA

combination improvements 2: SPATRY

15

combination improvements 3: YVCHWA

consensus: YVCHWASSATRY

activity at 0.01 μ g/ml